Psychological Monographs

General and Applied

lo. 457
1958

ol. 72
lo. 4

Louis M. Smith

The Concurrent Validity of Six Personality and Adjustment Tests for Children

By

Louis M. Smith

Graduate Institute of Education
Washington University, St. Louis

Price $1.00

Edited by Norman L. Munn

Published by The American Psychological Association, Inc.

Psychological Monographs:
General and Applied

Combining the *Applied Psychology Monographs* and the *Archives of Psychology* with the *Psychological Monographs*

Manuscripts and correspondence on editorial matters should be sent to the Editor. *Psychological Monographs* publishes comprehensive experimental investigations and programmatic studies which do not lend themselves to adequate presentation as journal articles. Major space is given to the author's original contribution; introductory and bibliographic materials, as well as statistical tables and graphs, must be kept within reasonable bounds. Tables, graphs, and appendix materials which deal with detail not essential to adequate presentation of the findings may be made available through the American Documentation Institute—for details of this procedure, see the APA *Publication Manual*. Preparation of manuscripts for publication as monographs should follow the procedure given in the APA *Publication Manual*. Publication in *Psychological Monographs* is free of cost to the author, except in cases where early publication is requested or author's alterations are made in galley proofs.

Correspondence on business matters should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D.C. Address changes must arrive by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee third-class forwarding postage.

Vol. 72, No. 4 Whole No. 457, 1958

Psychological Monographs: General and Applied

The Concurrent Validity of Six Personality and Adjustment Tests for Children

Louis M. Smith[1]

Graduate Institute of Education, Washington University, St. Louis

The present study investigates the validity of six personality and adjustment tests for children. In the terminology of the recent publication by the American Psychological Association Committee on Test Standards, the approach is one of "concurrent validity." The committee's report states:

> Concurrent validity is evaluated by showing how well test scores correspond to measures of concurrent criterion performance or status. Studies which determine whether a test discriminates between presently identifiable groups are concerned with concurrent validity (2, p. 14).

More specifically, this investigation indicates the relationship of the tests to a dual standard—teacher nominations of very well adjusted and very poorly adjusted children and peer nominations of classmates liked and classmates with whom one does not get along.

Three of the six tests are widely used instruments: California Test of Personality (52), Rosenzweig Picture-Frustration Study (43), and Rogers Test of Personality Adjustment (41, 42). Three are recently developed and as yet unpublished: How I Feel About Things, How Would You Finish It, and Things I Like To Do (4, 5, 6). All six are group instruments; they represent the structured test approach to adjustment testing; and they all yield a single score indicating level of adjustment.

In developing a rationale for an investigation such as this, several points must be considered. First, those persons who work with children have no feelings of doubt concerning the need for further evaluation of adjustment tests. A number of writers in Buros' several editions on tests and mental measurements make this point directly and by implication. In 1940, for instance, Louttit, in writing about the Rogers test, then 10 years old, described it ". . . with the exception of the time consuming method of scoring, have found it the most satisfactory instrument of personality measurement" (30, p. 94). Yet he cited only two references, including Rogers' original monograph, wherein the adequacy of the test was studied. More recently, Ellis' two reviews (15, 16) hardly mention a reference to this test's validity. Practically every author who discusses the California Test of Personality compliments the test authors on item writing, reliability of full test, adaptation for machine scoring, and other such points, but they consistently point up the lack of validity data for both the test as a whole and the individual items. Shaffer for example maintains, "It must be concluded that the validity of these questionnaires is entirely unestablished" (44, p. 56). Similar comments could easily be quoted concerning the Rosenzweig P-F Study.

Complementing this point of view, Kearny (27) demonstrates the important need felt by teachers. His opinionnaire study indicated that elementary school teachers in St. Paul are most eager for extended help in the area of guidance services. Integral to this are more adequate measures of social and emotional growth and development.

A third aspect of this rationale involves the considerable definitional confusion which exists

[1] This report is based on a dissertation submitted to the University of Minnesota in partial fulfillment of the requirements for the degree of Doctor of Philosophy, December 1955. The writer wishes to express his thanks to Mildred Templin and Ralph Berdie, his thesis advisers, for their helpful criticisms and suggestions.

in the variable being measured by these tests. The test titles and manual discussion refer to personality, personality adjustment, mental health, and group conformity, as well as a variety of other traits and symptoms. The present writer considers two concepts as crucial: adjustment and mental health. The definition of adjustment used in this investigation is the satisfactoriness of the relationship between the individual, a sixth-grade boy, and the environment, the sixth-grade elementary class. It is assumed that a continuum of poor to good adjustment exists and that each boy can be placed on this continuum. It is also assumed that the position on this continuum is reflected within the individual in terms of a "state" which ramifies into the individual's verbal behavior and is capable of being measured. "State," in this sense, refers to the individual's mental health or attitudes toward himself and others.

Operationally the variable which is assessed within the individual depends on the selection of the criterion groups. Teachers were asked to nominate boys who were "well adjusted" and boys who were "poorly adjusted" (see the defining paragraphs in the Appendix). Children in the sixth-grade classes were asked to nominate children they liked best and children with whom they did not get along. These operations indicate the meaning of "satisfactoriness of relationships" discussed above. The attitude structure of the boys chosen at each level becomes the definition of mental health continuum. The implication of these definitions as they relate to Havighurst's (21) use of the concept of developmental task, Symonds' (49) use of ego strength and Ackerson's (1) analysis of behavior problem types is given in Smith (45).

Another pertinent consideration involves the criterion. When, as in the present analysis, the criterion in an investigation utilizing predictive or concurrent validity is an intermediary one, it is necessary to indicate the further correlates of the criterion. By reviewing the large literature on sociometric choice one finds substantial support that children highly chosen as friends possess characteristics and behavior usually described as mentally healthy and well adjusted, and that children who are not chosen positively and who are rejected possess characteristics and behavior indicating maladjustment and mental ill health. Smith (45) has organized this literature into opinions (18, 19, 40, 56), pupil ratings (7, 8, 28, 32), adult ratings (9, 17, 25, 37), camp leader ratings (23, 35, 36), and psychiatric referrals (50). If a large part of the variance in sociometric acceptance can be assigned to personality and adjustment variables, and if the other major variables can be experimentally controlled or partialled out statistically, it is concluded that sociometric status is an acceptable criterion, or portion of a criterion, in the establishment of groups of poorly adjusted and highly adjusted children.

This investigation involves a second judgment, teacher nominations of pupils, as part of the criterion. Such a criterion technique in locating maladjustment in children contains many hazards. Wickman's study (55) and its replications and the many studies it influenced with their usual interpretations raise considerable doubt that classroom teachers and mental hygienists have any agreement as to the nature of maladjustment in pupils (26, 29, 34, 47, 48, 51). However, the present writer (45), after carefully reviewing this literature and relying heavily upon an early (and much neglected) but important article by Goodwin Watson (54), feels that many of these doubts are unjustified. The nature of directions given to teachers, differences in meaning of terms for different groups, lack of appreciation of chronicity of symptoms, and the ambiguity of the relation of the symptom to the total behavioral context contain the essence of this analysis.

When teachers select "behavior-problem" cases there is a strong tendency to nominate routine-disturbing, overtly hostile, and aggressive children. These children are below average in achievement, ability, and socioeconomic status. Shifting directions from problem cases to maladjusted children elicits a different kind of child, i.e., a wider variety of regressive and atypical personality problems (10, 12, 14, 31, 39, 46, 57).

To use teacher nominations of maladjusted pupils care must be taken to stress in the instructions long-range difficulty in adjustment, difficulties inherent in the child, and problems related to getting along with peers, rather than the child who is difficult to handle in class routines. With such instructions one avoids many of the differences in terminology connotations and the rating of symptoms detached from the individual. The child is seen as the unit and one evaluates his personal-social effectiveness in his environment.

The actual selection of tests to be analyzed in this project was made in the light of several criteria: (*a*) the tests must be amenable to group administration; (*b*) they must have some evidence of validity, past or potential; (*c*) they must be widely used; (*d*) in combination they must represent a variety of item types. The reasoning behind such requirements or standards requires some explication.

One of the original goals of this investigation is the measurement of adjustment in a classroom situation wherein screening rather than case-study analysis is important. Group administration is extremely important in this regard. That such a requirement constitutes a decided limitation is not denied. The mention of such instruments as the Bender Gestalt Test, the Thematic Apperception Test, and the Rorschach clearly indicates this. Criteria *b* and *c* actually were held in combination. Few personality and adjustment tests for children present much data evidencing strong validity. Some tests are currently being used widely with little regard for their underlying adequacy or validity. And finally, in the event of negative findings, criterion *d* might lend hypotheses for future research.

Methodology of the Study

An investigation analyzing the adequacy of six adjustment tests for children poses several methodological problems: (*a*) specification and analysis of the population from which the sample is drawn; (*b*) description of methods and procedures used in obtaining criterion groups at the different adjustment levels; (*c*) description of the sample in terms of age, parental occupational level, intelligence, and reading achievement; and (*d*) independent estimates of the validity of the criterion groups.

Population Sampled

All the children used in this study came from the public schools of St. Paul, Minnesota. Consequently, certain limitations occur in the generalizability of the findings: (*a*) the subjects live in an urban community; (*b*) all of the subjects attend public schools although approximately 40% of the children of elementary school age do not; (*c*) the public schools maintain classes for children with gross physical disability, both sensory and motor, and mental retardation; (*d*) other handicaps, such as reading disability and emotional disturbance, plus special abilities, such as those of gifted pupils, are handled in the regular classroom.[2]

The criterion used in the selection of participating schools from the total available was that the sixth grades must not contain fifth- or seventh-grade pupils and the classes must be self-contained, i.e., one teacher for all subjects. This leaves 29 schools, from a total of 66, containing 37 sixth-grade classes eligible for the investigation; the study includes all 37 classes. The 37 classes contain 653 boys. Several benefits accrue to the investigation when all the pupils have the same teacher for the whole day and all are in the sixth grade. The teachers know the pupils in several phases of the academic work and also have a chance to observe the pupils at the start and close of the day, during recess, and in a number of nonacademic situations. Similarly, the pupils have a wider variety of contacts among their peers. By restricting to one grade level, the learning situation has one less major schism.

Procedures in Selecting Pupils

One criterion for the selection of the subjects consists of teacher nominations. Each of the 37 teachers received a typewritten form[3] devised by the present investigator on which to make nominations of high and low adjustment status. The form contains descriptive paragraphs indicating the meaning of "very good adjustment" and "very poor adjustment" and space for writing in names of pupils who fit these descriptions. It also contains a third part for the names of pupils the teachers do not know well enough to nominate. Statements respecting anonymity of teachers and pupils and confidentiality of records are included in hopes of further validity.

The descriptive paragraphs indicating adjustment level reflect several considerations. A balance is struck between fairly general referents such as "well adjusted" and "emotionally disturbed" and fairly specific referents such as "doesn't lose temper easily" and "cries easily." Efforts are made to allow for a wide range of toleration for robust and spirited activity on the part of the adjusted group. Similarly efforts are made for difficulties other than the classroom disturber, i.e., "excessively shy and unhappy child" to fall in the maladjusted group. Finally, the description attempts to place the child's development on the long-range perspective of the significance for future adjustment.

The nominations by the teachers show some variability. First, the nominations for the "can-

[2] Administrative Offices, St. Paul Public Schools. Personal communication, 1955.

[3] This form is reproduced in the Appendix.

not say" category represent a problem, the degree to which teachers felt unsure about a pupil's adjustment status. Thirty-eight per cent of the teachers nominated two children and 24% nominated one child. Frequently the teachers used this with new pupils, but often with those with whom the teacher was not well enough acquainted to make a judgment. Logically such a procedure should eliminate a considerable number of cases of inadequate knowledge of pupils.

The teacher nominations contain a second type of variability. Ninety-two per cent nominated four or five very well adjusted boys, but only 68% nominated four or five very poorly adjusted boys in the class. Such percentages indicate clearly that the teachers recognized and/or were more willing to classify children at the high end of the adjustment continuum.

Sociometric nominations resolve into positive and negative choices, both of which were used. This choosing was accomplished by using another instruction form[1] devised by the investigator which was read to each class. First, the children designated "the classmates they liked best and got along with best." These names were placed on the top of the page. Second, on the bottom half of the paper, the pupils nominated or designated "the children with whom they and their classmates didn't get along." Although the children could nominate as many as they wished, it was suggested that they select four or five for each category. Even though the study involves only boys, it was decided that the entire class climate, both girls and boys, should be assessed and utilized in the selection of the criterion groups.

A tabulation of these positive and negative nominations was made. There is a high degree of variability in the number of nominations received by individual boys. The range is from 0 to 25 positive nominations and 0 to —26 negative nominations. The two lists were made into one ranking by counting all favorable nominations received +1 and all unfavorable nominations received —1. This gives a peer acceptance hierarchy. For example, one of the boys may have been nominated by each of twenty peers as one of several best friends. Three other classmates may have listed him as a classmate with whom they did not get along. Seven others might not have nominated him for either category. Consequently his score would be +17. Another boy may have been listed by four classmates as one of their best friends but listed by twelve others as one with whom they did not get along well and not nominated by fourteen. His score on the combined ranking would be —8.

[1] This form is reproduced in the Appendix.

There are several aspects of the sociometric peer nominations that should be noted. First, a wide variability from class to class occurs in the ratios of frequency of positive nomination to frequency of negative nomination. Table 1 presents these results. Here it is seen that the classes varied markedly in being positively and negatively oriented in nominations of boys. The trend to nominate more boys as good friends and well liked than to nominate them as boys with whom one does not get along is very clear. In only one class are there more negative than positive nominations. This predominance of positive feeling is also found to exist in the teacher nominations of pupils.

TABLE 1

DISTRIBUTION BY SCHOOL CLASSES OF THE RATIOS OF POSITIVE TO NEGATIVE CHOICES RECEIVED BY ALL BOYS IN POPULATION

Ratios of positive to negative choices received	*N*	%
2.30–	3	8
2.10–2.29	5	13
1.90–2.09	7	19
1.70–1.87	3	8
1.50–1.69	7	19
1.30–1.49	5	13
1.00–1.29	6	16
– .99	1	3
Total	37	99

When analyzed by individuals within a class, it is seen that the negative nominations, as well as the positive nominations, tend to cluster around a small number of pupils. The mental health effects of this both as cause and effect in maladjustment present an interesting problem beyond the scope of this discussion. For illustrative purposes a tabulated class record of peer choices is included in Table 2. This specific record contains a bias in being exceptionally clear-cut in differentiating the high and low acceptance of individual boys. Pupils B, D, and K, for instance, are operating in an entirely different social and emotional climate than are pupils A, I, and Q.

For all of the boys who ultimately became part of the 297 in the initial adjustment groups, using the combined teacher and pupil nomination techniques, the range in sociometric nominations is of interest. These data are presented in Table 3. Only a slight overlapping of the distributions occurs. Considering the range of positive to negative class nomination ratios, such clear sociometric differences in total distributions as presented in this table give an important indication of the method's success.

TABLE 2

ILLUSTRATIVE CLASS RECORD OF SOCIOMETRIC NOMINATIONS

Pupil	Positive nominations	Negative nominations	Combined nominations	Rank in class
A	4	14	−10	16
B	25	0	25	1
C	6	2	4	7
D	21	0	21	2
E	4	2	2	10.5
F	2	8	− 6	15
G	2	4	− 2	13
H	8	2	6	6
I	4	15	−11	17
J	5	2	3	8.5
K	17	0	17	3
L	3	0	3	8.5
M	3	4	− 1	12
N	4	2	2	10.5
O	11	0	11	4
P	1	6	− 5	14
Q	1	16	−15	18
R	11	2	9	5

The Combination Criterion

The literature indicates that there are other than adjustment correlates of each of the criterion techniques, therefore a combination of criteria should prove more advantageous than either used separately. Consequently the following procedures were followed. The teacher ratings already were conveniently in three categories—those nominated as well adjusted, those nominated as poorly adjusted, and a middle group of unnominated persons. No mention of the last group was made to the teacher; however, on a logical basis, those who remain after the top and the bottom extremes and the unknowns are eliminated would be in the average adjusted range. The composite accepted and rejected peer nomination hierarchy was broken arbitrarily into thirds.

TABLE 3

COMPARISON OF DISTRIBUTIONS OF THE COMBINED PEER NOMINATIONS RECEIVED FOR THREE ORIGINAL ADJUSTMENT GROUPS

Statistic	Adjustment groups		
	Poorly adjusted	Average adjusted	Well adjusted
R_u	2	7	25
Q_3	− 1	5	11
Mdn	− 5	2	9
Q_1	−10	1	7
R_l	−25	−1	3

The subjects selected then met these requirements: for the well adjusted group the pupil must be in the top third on sociometrics and on the teacher's list as very well adjusted. For the maladjusted group the pupil must be in the bottom third of the sociometrics and must be listed as maladjusted by the teacher. The average adjusted must be in the middle third of the sociometric rank and not be listed either high or low by the teacher. Those who did not meet these two criteria were considered doubtful as to adjustment status and were eliminated from the study.

To clarify the procedure a previous example is continued. The boy with a +17 score, for instance, is in the top third of the sociometric hierarchy. If his teacher has also selected him as well adjusted, then he falls in the well adjusted group. Another boy selected by the teacher as well as adjusted but falling in the middle group on the sociometrics would not be selected because of not meeting the dual standard. A similar procedure was used in the average adjusted and in the maladjusted groups. In general the teacher nominations agreed with the pupil nominations; however, there were some serious differences where a pupil was found to be well liked but not considered adjusted by the teacher. The reverse also occurred. A three-by-three contingency table with nine cells was constructed to determine the relationship between the two selection techniques. The contingency coefficient was .47.

The initial criterion groups chosen by these techniques require comment. This writer believes the techniques are not adequate for giving information about every child in the class. For instance there is the child who has been nominated in the well liked category by the children and placed by the teacher in the poorly adjusted category. For the present purposes he remains an enigma. He becomes one of the cases that lowers the correlation of each technique with the hypothetically pure rater of adjustment. He is not understood with the data at hand, and for this reason he is omitted from the present study. However, the fact that some pupils cannot be classified should in no way detract from the validity of the classification of those about whom both criteria agree.

Because all the tests studied are of the paper-and-pencil variety and require a minimum of fourth-grade reading skill, it was imperative that a final selection be made on reading achievement (before identifying the final three groups at different adjustment levels). The Gates Advanced Primary Reading Test, Part Two, Paragraph Reading, was selected for this purpose. Although

TABLE 4

COMPARISON OF THE THREE ADJUSTMENT GROUPS IN TERMS OF DIFFERENTIAL REDUCTION IN SAMPLE SIZE

Variable	Adjustment groups			
	Poorly adjusted	Average adjusted	Well adjusted	Total
Pupils initially selected by dual criteria	92	99	106	297
Pupils eliminated due to absences and moving	6	2	4	12
Pupils eliminated due to reading achievement below 4.0	23	15	2	40
Pupils finally tested	63	82	100	245
Total number of boys in the 37 classes				653

designed for lower grades, it discriminates quite adequately between grades 3.5 and 4.5. The classroom teachers gave this test to the pupils in the initial adjustment groups when the grade status was approximately 6.4. All of those pupils who scored 4.0 and above were scheduled for final testing with the six personality and adjustment tests.

A few additional pupils were lost due to moving and absences during the process of carrying out the investigation. The specific number of individuals lost is in Table 4.

In summary, the actual procedures used in the selection of the pupils were executed fairly simply. The teachers nominated up to five children as very well adjusted and very poorly adjusted. The pupils nominated classmates they liked as best friends and classmates with whom they did not get along. The two lists of nominations were combined to select children at three levels of adjustment. A reading test was administered to eliminate children from the criterion groups who were below fourth grade in reading achievement. A few children were lost from the criterion groups due to illness and changes in residence. The remaining children became the sample of pupils tested. These are the children on whom the results of the study are based and about whom the conclusions are drawn.

Test Administration Procedures

The actual administration of the six tests of personality adjustment occurred as follows. The tests were given in a random order except for the final few schools. Here a minimum rearrangement was carried out to force the tests into a pattern of no test ever being administered more than six times as the first test of the group, the second, or the third, and so forth. In this way no test had any benefits or handicaps over the others accruing from being administered at the beginning, middle, or end of the testing session. A short "break" was taken after the first hour to hour and a half.

The examiner read a large part of the tests even though the pupils had their own booklets. The suggestion for this procedure arose from the first group tested. The subjects, almost unanimously, favored such reading. This step speeded the test administration time considerably and decreased the work and potential fatigue of the pupils. It minimized the individual differences in working pace. In general the groups stayed pretty well together; however, the more able pupils often went ahead at their own rate on each test.

Since all subjects started each test at the same time, any pupils finishing early were permitted to draw on the back of the booklets while waiting for the others to finish. The children took wide advantage of this opportunity on the backs of the Rosenzweig P-F Study and the How Would You Finish It? test, for these were usually not read by the examiner. Incidentally the drawing on the booklets interested the children, occupied their free time, and added a considerable degree of control to the situation.

Although the effects of such concentrated testing are not determined, the following observations are noted. In general, the pupils were quite eager to be out of class for most of the morning. Most of them enjoyed the variety in the testing, and only when the California Test of Personality and the How I Feel About Things came in sequence did they become restive. Most of the children took the tests seriously although they would laugh or make humorous comments at individual items. Some of the pupils were quite flippant and difficult to control for the period. The break in the middle of the testing period proved relaxing. The most unhappy aspect occurred if a recess period was missed. At the close of the testing most of the pupils were about ready to go back to class; however, as they left they invariably would comment that if any more tests were to be taken they would like to be chosen.

Sample

With a description relative to the techniques for obtaining the subjects at the three adjust-

TABLE 5

COMPARISON OF FINAL THREE ADJUSTMENT GROUPS ON FOUR DESCRIPTIVE VARIABLES

Variable	Adjustment groups		
	Poorly adjusted $N = 63$	Average adjusted $N = 82$	Well adjusted $N = 100$
Age			
Mean	11.68	11.56	11.60
SD	.55	.55	.34
Intelligence			
Mean	97.13	102.61	107.91
SD	10.09	10.35	10.34
Reading Achievement			
Mean	5.61	6.10	6.48
SD	1.05	1.09	1.11
Parental Occupational Level			
Mean	2.48	2.34	2.12
SD	.71	.77	.80

ment levels in mind, further elaboration is required to better understand the sample studied. The variables considered are age, intelligence, reading achievement, and parental occupational level.

The age of the subjects as of December 31, 1954, is used for a constant reference. Since all the pupils were tested within a period of two months such a procedure seems feasible. The average age for each adjustment group is slightly over 11½. The simple technique of subtracting five gives the grade status in which the pupils should be. This is halfway through the sixth grade. The standard deviations, although significantly different for the three groups, range from about one-third to one-half year and indicate, for practical purposes, that the groups are homogeneous. This reflects a strong age-grade promotion policy within the St. Paul Public Schools.

Intelligence test scores on the Otis Quick Scoring Test of Mental Ability (38) from the city-wide testing program were available for most of the pupils. Some of the pupils had been given the individual Stanford-Binet and Wechsler Intelligence Test for Children. If the group tests were not available, these were used. Some half dozen pupils with neither record available were given the Ammons Picture Vocabulary Test (3) as a rapid screening measure of intelligence. As indicated in Tables 5 and 6, the groups are significantly different in intelligence.

The results from the Gates Advanced Primary, Paragraph Reading Test (20) originally used in screening the pupils also were used as a general reading achievement measure. Because of the limited ceiling on this test, any differences occurring probably would be a minimal reflection of differences in actual ability. Despite the differential elimination of a number of poor readers and the lack of ceiling on the test, the adjustment groups are very significantly different in reading achievement. Pupils nominated by their teachers as very well adjusted and nominated by their peers as best friend are achieving pupils as compared to pupils comprising the average and poorly adjusted groups.

The Minnesota Scale for Parental Occupations (53) was used to classify the level of parental occupation. School records supplemented by teacher and pupil information were the sources of the data. To obtain categories with large enough numbers to handle, the seven groupings were combined into three. Classes I and II became Group One, Classes III and IV became Group Two, and Classes V, VI, and VII

TABLE 6

ANALYSIS OF VARIANCE OF THE FINAL THREE ADJUSTMENT GROUPS ON FOUR DESCRIPTIVE VARIABLES

Variable	Source of variation	*df*	Mean square	F^a	*p*
Age	Between groups	2	.27	1.40	$> .05$
	Within groups	242	.18		
	Total	244			
Intelligence	Between groups	2	2281.09	21.58	$< .01$
	Within groups	242	105.75		
	Total	244			
Reading Achievement	Between groups	2	20.68	17.46	$< .01$
	Within groups	242	1.18		
	Total	244			
Parental Occupational Level	Between groups	2	2.64	4.42	$< .05 > .01$
	Within groups	242	.60		
	Total	244			

[a] The tests of homogeneity of variances to determine the tenability of the technique's assumptions are recorded in the Appendix.

became Group Three. Group differences in level of parental occupation show a positive relation between socioeconomic status and good adjustment. Such findings are consistent with the intelligence and achievement differences.

In summary there were significant differences for consistently higher mean scores moving from poor to good adjustment on intelligence, reading achievement, and level of parental occupation. No differences existed in mean ages of the groups. Table 4, presented earlier, gives the results of selection techniques in terms of sample size findings.

Independent Estimate of Criterion Validity

Earlier, an assumption was stated that three levels of general adjustment could be established through a combined technique of teacher and pupil nominations. Reference to literature supporting this contention has already been made. Four independent estimates of the validity of this assumption clarify this point. One of these, pupils who are currently serving on the school police patrol, is an estimate of relative honor and responsibility, an attribute of adjustment. Referrals to the school social workers, referrals to the community child guidance clinic, and arrests and records at Juvenile Police all indicate children with problems and difficulties of adjustment. Although complete independence of each validity estimate is desirable, this goal has not been reached. Principals and teachers select pupils for the school police and make referrals to the school social workers. Referrals to the community guidance clinic originate from parents and community agencies including the school social workers and the juvenile court. Arrests and records at the juvenile division are strictly a police matter.

These independent data estimate the adequacy of the combination criterion, pupil and teacher nominations; consequently, the initial adjustment groups totaling 297 pupils comprise the sample discussed rather than the final tested sample of 245 subjects. This larger group also is advantageous because the frequency of some of the behaviors is small.

Table 7 presents the data relative to membership on the school police patrol.[5] In this case the numbers in the totals are less than in the criterion groups because only 17 of the 29 schools have school patrols in which sixth-grade pupils are allowed to participate. A clear relationship occurs between level of adjustment and being on the force. The extreme categories of

TABLE 7

COMPARISON OF CURRENT PARTICIPATION ON SCHOOL POLICE FORCE BY PUPILS IN THE THREE ADJUSTMENT GROUPS

Level of participation	Adjustment groups					
	Poorly adjusted		Average adjusted		Well adjusted	
	N	%	*N*	%	*N*	%
Patrol officers	0	0	0	0	21	32
Patrol members	23	42	44	70	36	56
Not on patrol	32	58	19	30	8	12
Total	55	100	63	100	65	100

the table indicate this most clearly. Thirty-two per cent of the well adjusted group are officers on the force, while none of the average or poorly adjusted group are officers. Fifty-eight per cent of the poorly adjusted group did not make the force as compared to 12% of the well adjusted group. Two further comments are in order. A number of the members of the poorly adjusted group who are school police came from just a few schools, where practically the whole class is on the force. Second, a school police position is occasionally used by these principals and teachers as a motivational technique or perhaps as an "evidence of faith" to bolster the self-esteem and to aid in the social integration of pupils with adjustment difficulties.

The visiting teacher or school social worker is a part of the special services program available to the school personnel. Her job function lies primarily in problems of emotional disturbance, social maladjustment, and truancy. In general, the casework technique of interviewing children and parents, consulting with teachers and other school personnel, and making referrals to community agencies constitute their approach to the problem of maladjustment.

With such functions and procedures in mind, the data in Table 8 are presented (see footnote

TABLE 8

COMPARISON OF REFERRALS OF CHILDREN IN THREE ADJUSTMENT GROUPS TO THE SCHOOL SOCIAL WORKER

Time of referral	Number in adjustment groups[a]		
	Poorly adjusted	Average adjusted	Well adjusted
September 1954–March 1955	15	3	0
January 1953–June 1954	23	2	1
Total	38	5	1

[5] Office of the School Police, Department of Public Safety, St. Paul, Minnesota. Personal communication, 1955.

[a] These figures do not indicate number of different individuals, for a few of the children were referred both years.

2 preceding). During the current school year 18 children in the criterion groups were referred to the visiting teacher. Such referral is usually initiated by the teacher to the principal and then to the visiting teacher. This in and of itself would not give adequate support to the contention of adequacy of the criterion groups, for the teachers, as was described, contributed to the selection of the latter groups. However, the referrals for the year and a half preceding the present study indicated that there were other teachers who considered the children, or the situation, as needing outside help. And incidentally such data over a period of time show that for a number of the children described as poorly adjusted the problems are not transient but of a long-standing nature. Although not reported, data from earlier years indicate several were noted even then. There were also a number whose status has changed.

Twenty-five boys in the 57 classrooms were the total referrals to the visiting teachers during the present year. Of the seven not included in the study three were in the teachers' median adjusted group for they were not nominated by the teacher as very well or very poorly adjusted —but all three were in the bottom third on the sociometric peer nominations. Four children were nominated by the teacher as poorly adjusted. Of these there were two in the middle sociometric group. They bordered on the lower end of this group. The other two presented a serious discrepancy for on the peer nominations they were in the top third. Consequently, of the 25 referrals only two are seriously ambiguous in terms of the present analysis.

The community of St. Paul has a privately endowed child guidance clinic specializing in treatment of children with problems of adjustment. Table 9 presents the data concerning referrals and treatment at this clinic.[6] Of the current cases all but one are in the poorly adjusted group. Although the absolute numbers are not large, the relative differences in proportions of recently referred cases and cases in treatment is of interest. For the latter, 6 cases in 92 is only a percentage of 6.52; however, when one combines the average and well adjusted groups of children, only 1 case in 205 occurs. This is 0.487 per cent. The ratio is over 13 to 1 in favor of more referrals in the lower group.

The City of St. Paul, like many other urban communities, has a juvenile division within the police department.[7] All arrests of persons 18 years of age and younger are channeled through

TABLE 9

CHILDREN IN THE CRITERION GROUPS WHO HAVE BEEN REFERRED AND TREATED AT WILDER CHILD GUIDANCE CLINIC

Type of contact	Number in adjustment groups		
	Poorly adjusted	Average adjusted	Well adjusted
Current cases, January 1955	6	1	0
Recent referrals, January 1953–January 1955 (including current cases)	7	2	0
Treatment cases closed[a]	4	4	1
Referred but withdrawn	3	4	1

[a] Most of these case contacts occurred prior to 1953.

the juvenile division. Most children who are arrested come in with their parents for a conference. Usually the child is warned, restitution is made if there is property damage, and no further action is taken; however, some of the cases are taken to court and probation or commitment occurs.

The crimes for which children in the criterion groups were booked included such incidents as shoplifting a 10-cent item from a 5-and-10-cent variety store, property damage from several dollars to over a thousand dollars, automobile theft, arson, window peeping, running away from home, and sodomy. There were another half dozen categories of delinquencies so that practically the entire range of legal difficulties a child under 12 years of age could commit are included. That some of the crimes are insignificant from an adjustment point of view is not denied; however, one would expect a relatively lower frequency of arrests in the well adjusted and average adjusted groups as compared to the poorly adjusted group.

Almost 22% of the children in the poorly adjusted criterion group have been arrested at least once in the last two years. This is contrasted with 5% to 6% of the median and well adjusted groups. Arrests prior to this time also show a relationship between arrest frequency and difficulty in adjustment. For the arrested children the modal number is only one time arrested, but the chronicity of legal difficulties lies heavily within the poor adjustment group.

Although not tabled, a final set of facts is appropriate. This concerns the disposition of the cases. As mentioned earlier, parent conferences, restitutions, and warnings are the usual procedures. Only five individuals committed crimes serious enough by police standards to warrant referral to juvenile court for review by a judge. All five of the children are in the poor adjustment group. All five were placed on probation. Three children, again all in the poor adjustment group, were referred for psychiatric

[6] Thorsen, D. Personal communication, 1955.

[7] Roberts, J. Juvenile Division, St. Paul Police Department, Bureau of Public Safety. Personal communication, 1955.

TABLE 10

COMPARATIVE PERCENTAGES OF ARRESTED CHILDREN IN THE THREE CRITERION GROUPS ACCORDING TO DATES OF MOST RECENT ARREST

Date	Adjustment groups[a]					
	Poorly adjusted		Average adjusted		Well adjusted	
	N	% in Group	*N*	% in Group	*N*	% in Group
January 1953–March 1955	20	21.74	5	5.05	6	5.66
Prior to December 1952	6	6.52	4	4.04	2	1.89
Total	26	28.26	9	9.09	8	7.55

[a] *N* in adjustment groups is as follows: poorly adjusted 92; average adjusted, 99; well adjusted, 106.

diagnosis and treatment. Two are included in the statistics related to the Child Guidance Clinic and one received private psychiatric care. The precipitating incident in one case was running away from home and in the other it was sodomy.

In summary, the independent validity data consistently support the rationale of the selection of the criterion groups. There was a direct relationship between being on the school police force and adjustment level. All the police officers were from the well adjusted group. Decidedly more children in the poorly adjusted group were referred to the visiting teachers both this year and last; more have been arrested at least once and more have been arrested more than once; finally more were under psychiatric care in the community guidance clinic.

RESULTS OF THE INVESTIGATION

The analysis of the findings follows, in general, a simple organization. Descriptive statistics for each test are presented. The results of each test are analyzed by analysis of variance indicating statistical significance of mean differences, by analysis of covariance indicating mean differences when intelligence, parental occupation, and reading achievement are controlled and by percentage of overlap indicating amount of overlap of distributions. A brief analysis of the relationships among the tests is also included.

TABLE 11

COMPARISON OF TOTAL NUMBER OF ARRESTS OF CHILDREN IN THE THREE CRITERION GROUPS

Frequency of arrests	Adjustment groups		
	Poorly adjusted	Average adjusted	Well adjusted
1	17	5	8
2	4	4	0
3	2	0	0
4	1	0	0
5	1	0	0
6	0	0	0
7	0	0	0
8	1	0	0
Total *N* of arrests	48	13	8

California Test of Personality

The California Test of Personality contains two subtests, Personal Adjustment (PA) and Social Adjustment (SA). Results of the two subtests sum to give a single score indicating an individual's total adjustment status or level. Tables 12 and 13 contain the descriptive statistics—sample size, mean, and standard deviation and the analysis of variance results. The *F* tests in Table 13 are highly significant; the null hypotheses are rejected. In short, it can be concluded that this test does differentiate statistically between levels of adjustment for sixth-grade school children. Theoretically this

TABLE 12

DESCRIPTIVE STATISTICS FROM THE CALIFORNIA TEST OF PERSONALITY

Aspect of the test	Adjustment groups		
	Poorly adjusted *N* = 63	Average adjusted *N* = 82	Well adjusted *N* = 100
Total score			
Mean	96.35	106.93	111.96
SD	18.72	16.89	16.26
Personal adjustment			
Mean	47.76	53.37	55.95
SD	10.71	9.27	9.27
Social adjustment			
Mean	48.59	53.57	56.01
SD	9.56	9.24	8.10

TABLE 13

ANALYSIS OF VARIANCE OF THE CALIFORNIA TEST OF PERSONALITY

Aspect of the test	Source of variation	*df*	Mean square	*F*	*p*
Total score	Between groups	2	4,737.45	16.37	<.01
	Within groups	242	289.44		
	Total	244			
Personal adjustment	Between groups	2	1,304.85	13.04	<.01
	Within groups	242	93.53		
	Total	244			
Social adjustment	Between groups	2	1,019.84	12.02	<.01
	Within groups	242	78.92		
	Total	244			

is important in that it indicates the groups can be differentiated through their verbal behavior on a paper-and-pencil test.

The use of the analysis of covariance technique basically indicates similar results of mean differences; however, when pretest differences in intelligence, parental occupation, and reading level are equated, the *F* values consistently drop. Table 14 presents these findings.

It is interesting to note also that the mean values of the Personal Adjustment and Social Adjustment subtests vary less than a point within each level of adjustment. In terms of the test author's description of the test, there is no difference in personal and social adjustment within each group. The product-moment correlations between the subtests for each level of adjustment are .75 for the well adjusted group, .66 for the average adjusted group and .70 for the poorly adjusted group when not corrected for unreliability. Such analysis indicates that each subtest measures to a considerable degree the same thing.

Table 15, presenting percentage of overlap, gives an indication of the similarity of the distributions and the practical or psychological significance of the results. Within each level of adjustment the range in raw score values covers the complete scale. Such findings are one indication of the tenuousness in establishing cutoff scores to enhance prediction or judgment-making in regard to the individual case. The same conclusion occurs in reading Table 15 in terms of false positives and negatives. For in-

TABLE 14

ANALYSIS OF COVARIANCE OF CALIFORNIA TEST OF PERSONALITY (TOTAL SCORE)

Variable held constant	Source of variation	*df*	Mean square	*F*	*p*
Intelligence	Total	243			
	Within groups	241	288.45	9.80	<.01
	Adjusted mean	2	2828.17		
Reading achievement	Total	243			
	Within groups	241	293.18	12.04	<.01
	Adjusted mean	2	3530.91		
Parental occupational level	Total	243			
	Within groups	241	268.79	12.52	<.01
	Adjusted mean	2	3365.49		

TABLE 15

PERCENTAGES OF EACH ADJUSTMENT GROUP EXCEEDING THE MEAN OF THE OTHER GROUPS ON THE TOTAL SCORES OF THE CALIFORNIA TEST OF PERSONALITY

Adjustment group	Number and percentage above the mean					
	Poorly adjusted		Average adjusted		Well adjusted	
	N	%	*N*	%	*N*	%
Poorly adjusted			20	32	14	22
Average adjusted	60	73			32	30
Well adjusted	85	85	74	74		

stance, by using the average score of the poorly adjusted group as the cutoff point, one can screen 50% of the true positives, the maladjusted children. At the same time, one screens 27% of the average adjusted group and 15% of the well adjusted group as false positive. To improve the prediction of true positives one includes a growing percentage of false positives.

How Would You Finish It?

The How Would You Finish It? test is a 10-item sentence completion blank. It is scored on the affectivity of the response—positive, negative, or neutral—made by the subject to the stimulus words. The analysis of variance technique indicates that the groups differ statistically in the appropriate direction. The probability of such differences arising by chance is greater than .01 and less than .05. These data are in Table 16. The analysis of covariance, in equating intelligence scores of the groups, raises the value of *F* slightly; however, when initial differences in reading achievement and parental occupation are eliminated, the *F* value drops below the .05 level of significance. One concludes that the statistical significance of mean differences between groups on this test is ambiguous.

The distributions of the groups on this

TABLE 16

DESCRIPTIVE STATISTICS FOR THE HOW-WOULD-YOU-FINISH-IT TEST

Statistic	Adjustment groups		
	Poorly adjusted *N* = 63	Average adjusted *N* = 82	Well adjusted *N* = 100
Mean	15.05	15.20	16.00
SD	2.73	2.64	2.38

test show that each group has at least one member scoring as low as 9 and each has more than one individual scoring at the maximum of 20. The mode is 18 in all three groups. The distributions are negatively skewed. A tabular presentation of overlap percentages indicates a cutoff score obtaining 50% true positives will include 49% false positives from the average adjusted group and 39% false positives from the well adjusted group. For all practical purposes the distributions are identical.

How I Feel About Things

The How I Feel About Things test consists of an adjustment subtest and a combined adjustment-maturity subtest. The sum of these gives the full scale score. Both of the subtests and the full scale discriminate between the groups better than chance. All three *F* values indicating mean differences are highly significant. When the covariance technique is used to account for the initial differences in intelligence, reading achievement, and parental occupation, the *F* value of 11.33 drops to 4.93, 7.08, and 7.93 respectively. In short, the group differences would be less if an experimental control of these variables had been undertaken. These results are in Tables 17, 18 and 19.

The correlation coefficients of the subtests within each group are .65, .48 and

TABLE 17

DESCRIPTIVE STATISTICS FROM THE HOW-I-FEEL-ABOUT-THINGS TEST

Aspect of the test	Adjustments groups		
	Poorly adjusted $N=63$	Average adjusted $N=82$	Well adjusted $N=100$
Full scale			
Mean	27.54	35.32	39.68
SD	16.84	16.11	14.98
Adjustment subtest			
Mean	14.51	18.29	20.76
SD	8.08	7.93	7.48
Adjustment-maturity subtest			
Mean	13.03	16.93	18.92
SD	10.40	10.69	9.62

.50 respectively for the well adjusted, average adjusted, and poorly adjusted groups. The values indicate that the subtests are only partially independent of each other. In comparison, the subtests of the California Test of Personality are more similar than the subtests of the How I Feel About Things Test. These findings are noted because of the basic similarity in kinds of items within the two tests.

TABLE 20

PERCENTAGES OF EACH ADJUSTMENT GROUP EXCEEDING THE MEAN OF THE OTHER GROUPS ON THE TOTAL SCORES OF THE HOW-I-FEEL-ABOUT-THINGS TEST

Adjustment group	Number and percentage above the mean					
	Poorly adjusted		Average adjusted		Well adjusted	
	N	%	*N*	%	*N*	%
Poorly adjusted			24	38	18	29
Average adjusted	60	73			36	44
Well adjusted	82	82	70	70		

Table 20 indicates the significance of the test for individual prediction. Once again the wide range of scores in each adjustment level and extensive overlap

TABLE 18

ANALYSIS OF VARIANCE FOR THE HOW-I-FEEL-ABOUT-THINGS TEST

Aspect of the test	Source of variation	*df*	Mean square	*F*	*p*
Full scale	Between groups	2	2849.75	11.33	<.01
	Within groups	242	251.45		
	Total	244			
Adjustment	Between groups	2	755.46	12.44	<.01
	Within groups	242	60.75		
	Total	244			
Adjustment-maturity	Between groups	2	672.21	6.46	<.01
	Within groups	242	103.98		
	Total	244			

TABLE 19

ANALYSIS OF COVARIANCE OF THE HOW-I-FEEL-ABOUT-THINGS TEST

Variable held constant	Source of variation	*df*	Mean square	*F*	*p*
Intelligence	Total	243			
	Within groups	241	230.91	4.93	<.01
	Adjusted means	2	1182.92		
Reading achievement	Total	243			
	Within groups	241	248.30	7.08	<.01
	Adjusted means	2	1757.08		
Parental occupational level	Total	243			
	Within groups	241	225.43	7.97	<.01
	Adjusted means	2	1797.14		

make it impossible to ascertain adjustment levels for the individual pupil with this test. However, the continued finding of statistically highly significant differences in mean scores is of theoretical importance even though the test as such has more limited practical use.

Things I Like To Do

This is a test of positive and negative attitudes toward many activities in the child's day-to-day environment. It is scored in terms of total likes. The means and standard deviations are quite similar. The analysis of variance indicates there is no mean difference beyond chance in the response of the three groups to the test. The data are in Table 21. Application of the analysis of covariance indicates slightly higher F values; however, only in the instance of the correction for intellectual differences does the F indicate differences great enough to reject the hypothesis of no difference. Graphically the distributions overlap almost completely with the lowest score actually by a member of the well adjusted group, and the highest scores by members of the average and poorly adjusted groups. One must conclude that for purposes of selection, individual guidance and prediction, and research this test is not a valid indicator of adjustment status.

TABLE 21

DESCRIPTIVE STATISTICS FOR THE THINGS-I-LIKE-TO-DO TEST

Statistic	Adjustment groups		
	Poorly adjusted $N=63$	Average adjusted $N=82$	Well adjusted $N=100$
Mean	24.30	24.65	25.79
SD	5.37	5.13	5.15

TABLE 22

DESCRIPTIVE STATISTICS FOR THE ROGERS TEST OF PERSONALITY ADJUSTMENT

Aspect of the test	Adjustment groups		
	Poorly adjusted $N=61$	Average adjusted $N=81$	Well adjusted $N=99$
Total score			
Mean	39.18	35.95	32.01
SD	9.43	8.25	7.46
Personal Inferiority			
Mean	12.33	11.57	9.61
SD	4.32	4.48	3.41
Social Maladjustment			
Mean	15.13	13.32	12.82
SD	5.18	4.46	4.32
Family Maladjustment			
Mean	8.62	8.26	7.20
SD	3.46	3.61	3.13
Daydreaming			
Mean	3.10	2.80	2.37
SD	2.35	2.14	2.26

Rogers Test of Personality Adjustment

In spite of the precautions taken by the investigator, the Rogers test was responded to inappropriately in four cases; consequently, four fewer cases are in the total group. Two are lost from the poorly adjusted group, and one each from the average and well adjusted groups.

The analysis of variance technique indicates significant differences between groups on total score. On this test high scores indicate maladjustment; this is contrary to the other tests. Two of the scales have F values with a probability of less than .01. The most discriminating scale is labeled Personal Inferiority[8]; second is the Social Maladjustment Scale. The Family Maladjustment Scale has an F value greater than .01 but less than .05. The Daydreaming Scale has an F value greater than .05. All of the scale differences are in the direction expected by the test author. The analysis of covariance results are similar to results

[8] This value is spuriously high because the assumption of homogeneity of the variances is not tenable. The results of the Bartlett test are recorded in the Appendix. This is the only instance in this section on results in which the assumption of homogeneity of variances was not met.

TABLE 23

ANALYSIS OF VARIANCE FOR THE ROGERS TEST OF PERSONALITY ADJUSTMENT

Aspect of the test	Source of variation	*df*	Mean square	*F*	*p*
Total score	Between groups	2	1009.41	14.78	<.01
	Within groups	238	63.31		
	Total	240			
Personal inferiority	Between groups	2	162.77	9.91	<.01
	Within groups	238	16.42		
	Total	240			
Social maladjustment	Between groups	2	104.07	4.92	<.01
	Within groups	238	21.16		
	Total	240			
Family maladjustment	Between groups	2	45.25	3.91	<.05>.01
	Within groups	238	11.57		
	Total	240			
Daydreaming	Between groups	2	10.53	2.08	<.05
	Within groups	238	5.06		
	Total	240			

TABLE 24

ANALYSIS OF COVARIANCE FOR THE ROGERS TEST OF PERSONALITY ADJUSTMENT

Variable held constant	Source of variation	*df*	Mean square	*F*	*p*
Intelligence	Total	239			
	Within group	237	65.05	8.83	<.01
	Adjusted means	2	575.52		
Reading achievement	Total	239			
	Within group	237	68.57	12.54	<.01
	Adjusted means	2	859.68		
Parental occupational level	Total	239			
	Within groups	237	68.27	13.26	<.01
	Adjusted means	2	905.21		

obtained from most of the other tests. When initial differences in intelligence, reading achievement, and parental occupation are accounted for, then the values of *F* computed from total test scores show a consistent drop. In this test, however, they all remain statistically significant. The total score distributions as given in the tabular presentation of the data shows the high degree of overlap and the problems involved in using this test for individual prediction.

TABLE 25

PERCENTAGES OF EACH ADJUSTMENT GROUP BELOW THE MEAN OF THE OTHER GROUPS ON THE TOTAL SCORE OF THE ROGERS TEST OF PERSONALITY ADJUSTMENT

Adjustment group	Number and percentage below the mean					
	Poorly adjusted		Average adjusted		Well adjusted	
	N	%	*N*	%	*N*	%
Poorly adjusted			24	39	16	26
Average adjusted	57	70			31	38
Well adjusted	81	82	68	69		

An analysis of the relationship between the subtests or scales of the Rogers Test indicates approximately zero correlations. Within each adjustment group as described by the criterion the individual who scores high on one scale will not

TABLE 26

CORRELATION MATRIX OF THE FOUR SCALES OF THE ROGERS TEST OF PERSONALITY ADJUSTMENT FOR EACH ADJUSTMENT GROUP

Variable	Personal Inferiority	Social Maladjustment	Family Maladjustment	Day dreaming
Personal Inferiority				
W[a]	—	.04	.19	.02
A		.04	.23	.06
P		.16	.24	.01
Social Maladjustment				
W		—	−.04	.13
A			.15	.11
P			.12	.30
Family Maladjustment				
W				.17
A			—	−.02
P				.08

[a] Abbreviations for *Well* Adjusted, *Average* Adjusted, and *Poorly* Adjusted Groups reading from top to bottom.

necessarily score high or low on any of the three other scales. Such lack of relationship raises serious question concerning the feasibility of arithmetically combining scores on each part into a single or common total score. The results contrast vividly with the previously presented results of the interrelationships of the California Test of Personality and the How I Feel About Things Test. The parts of these tests showed subtest correlations of approximately .7 and .5 for the three adjustment groups. This lack of relationship in the Rogers test merits more examination.

Rosenzweig Picture-Frustration Study

The Rosenzweig P-F Study contains several novelties of scoring and tabulating. Consequently the statistical analysis of the results only partially follows the pattern of the five previous tests. The Group Conformity Rating (GCR) is a single quantitative score distributed approximately normally for each group. Its meaning is similar to the scores on the previous tests in that the lower the score the more atypical the result and interpretation. The previous pattern of descriptive statistics, analysis of variance, analysis of covariance, and percentage of overlap, is followed with the GCR.

Logically, the three scores indicating direction of aggression, the three scores indicating type of reaction, and the nine scores from combining the above would not be expected to relate linearly to adjustment status. For example, an exceedingly high intropunitive score showing excessive guilt feelings might be as typical of a maladjusted child as would an exceedingly low score showing complete lack of guilt. Similarly, a balance of responses to frustration might be more typical of the well adjusted boy. No such formal assumption is made, however, for in the last analysis the empirical findings must indicate the significance of the scores in terms of adjustment. Because of such considerations, the chi-square technique is used to ascertain significant differences between the groups for these variables. No analysis is made for super ego responses because of the infrequent stimulus situations. Only one trend score is analyzed, the change in GCR scores from the first to the second half of the test.

There are seven incomplete or illegible booklets despite the investigator's precautions. Three are from the poorly adjusted group, three are from the average group, and one is from the well adjusted group. The total sample used for statistical analysis of the Rosenzweig P-F Study numbers 238.

The results of the Group Conformity Rating follow the findings of the previous tests: the well adjusted group has the highest mean score, the groups are significantly different, and the score is related to adjustment status. The analysis of covariance indicates a slight drop in the F values when intelligence and

TABLE 27

DESCRIPTIVE STATISTICS FOR THE GCR OF THE ROSENZWEIG P-F STUDY

	Adjustment groups		
	Poorly adjusted	Average adjusted	Well adjusted
Mean	13.36	14.18	14.77
SD	2.37	2.09	2.08

TABLE 28

ANALYSIS OF VARIANCE OF THE GCR OF THE ROSENZWEIG P-F STUDY

Source of variation	*df*	Mean square	*F*	*p*
Between groups	2	27.64	5.88	<.01
Within groups	235	7.07		
Total	237			

reading achievement are controlled. The *F* value rises slightly in accounting for differences in parental occupation. A graphic analysis also indicates results similar to findings on the previous tests. Each group is distributed over the same range and has slightly different mean values, but the overlapping is very great. Tables 27, 28, 29, and 30 present these findings.

The GCR trend changes are analyzed in the following manner. A score for each half of the test was tabulated in making up the total score. For each group it was possible then to determine if the value on the second half is higher, lower, or the same compared to the value on the first half. A nine-entry chi-square table was constructed. The X^2 of 6.24 with four degrees of freedom has a probability greater than .05. One concludes that there is no difference in trends on the GCR. The greatest contribution to the chi-square value came from the small frequencies in the no-difference categories. The majority of each adjustment group tended to obtain higher ratings on the latter half of the test (see Table 31).

As discussed earlier, the scores on direction of aggression and type of reaction are grouped to account for differences in nature of the distributions as well as differences in mean scores. The results of the chi-square analyses appear

TABLE 30

PERCENTAGES OF EACH ADJUSTMENT GROUP EXCEEDING THE MEAN OF THE OTHER GROUPS ON THE GCR OF THE ROSENZWEIG P-F STUDY

Adjustment group	Number and percentage above the mean					
	Poorly adjusted		Average adjusted		Well adjusted	
	N	%	*N*	%	*N*	%
Poorly adjusted			23	38	21	35
Average adjusted	54	68			37	47
Well adjusted	80	81	60	61		

TABLE 29

ANALYSIS OF COVARIANCE OF THE GCR OF THE ROSENZWEIG P-F STUDY

Variable held constant	Source of variation	*df*	Mean square	*F*	*p*
Intelligence	Total	236			
	Within groups	234	4.72	4.78	<.01
	Adjusted means	2	22.59		
Reading achievement	Total	236			
	Within groups	234	4.72	5.73	<.01
	Adjusted means	2	27.03		
Parental occupational level	Total	236			
	Within groups	234	4.71	6.25	<.01
	Adjusted means	2	29.42		

TABLE 31

CHI-SQUARE TABLE FOR TREND CHANGES FROM THE FIRST TO THE LAST HALF OF THE TEST, IN GROUP CONFORMITY RATING SCORES OF THE ROSENZWEIG P-F STUDY

Trend changes	Adjustment group	Frequency		
		o	e	$(o-e)^2/e$
Higher GCR in second half	Poorly adjusted	34	34	.00
	Average adjusted	47	45	.09
	Well adjusted	55	57	.07
Lower GCR in second half	Poorly adjusted	22	20	.20
	Average adjusted	29	27	.15
	Well adjusted	30	34	.37
No change	Poorly adjusted	4	4	.20
	Average adjusted	3	7	2.28
	Well adjusted	14	9	2.78
Total		238	238	6.24

TABLE 32

CHI-SQUARE VALUES FOR 15 VARIABLES FROM THE ROSENZWEIG P-F STUDY

Variable	χ^2	*df*		*p*
Extrapunitive	6.91	8		>.05
Intropunitive	13.12	6	<.05	>.01
Impunitive	14.28	6	<.05	>.01
Obstacle Dominance	6.20	8		>.05
Ego Defensiveness	8.33	8		>.05
Need Persistence	7.48	6		>.05
E′	8.31	8		>.05
I′	1.92	4		>.05
M′	11.54	8		>.05
E	13.68	6	<.05	>.01
I	13.35	8		>.05
M	13.80	10		>.05
e	9.55	8		>.05
i	4.30	4		>.05
m	3.69	8		>.05

in Table 32. Three of the 15 variables are significant at the .05 level, none at the .01 level.

On the significant intropunitive variable, the distribution of the poorly adjusted group has a wider range, is more platykurtic, and tends to have more scores toward the low end than the average or well adjusted groups. High intropunitive scores are equally common to all three groups. The same type of relationship occurs for the significant impunitive variable. The distribution of the poorly adjusted group is platykurtic, has a wide range, and contains more scores at the low end of the distribution. The average adjusted group is considerably more leptokurtic. The extrapunitive-ego defensive variable, *E*, shows significant differences mainly because there is an excess of high scores in the poorly adjusted group and a minimum of high scores in the well adjusted group.

The same general cautiousness in interpretation of the results of the Rosenzweig P-F Study is required as in the other tests administered. The GCR scores show statistically significant different means for the three groups and the distributions overlap highly. Only 3 of the 15 chi-square values, regarding direction of aggression and type of reaction, are significant and in these 3, as well as in the other 12, the amount of overlap of the distributions is so great that individual prediction relative to adjustment is hazardous. For example, extremely high scores or low scores may indicate the dynamics of the individual in responding to frustrating situations; however, such scores taken singly tell you absolutely nothing about an individual's school adjustment status in 12 of the 15 variables and very little in regard to the other 3.

Relationship Among the Tests

Because of the manner in which the three groups were selected it is not possible to combine the three adjustment groups into a single group and compute the intercorrelations for the six tests. Consequently the relationships between the tests are computed within each adjustment group. For example, the scores on the California Test for the individuals in the well adjusted group are correlated with the scores on the How I Feel About Things Test. In the same manner the relationships between the same tests for the average and poorly adjusted groups are computed.

TABLE 33

CORRELATION MATRIX OF THE SIX ADJUSTMENT TESTS FOR EACH ADJUSTMENT GROUP

Test	HIF	Likes	S-C	Rogers[b]	GCR
California Test of Personality					
W[a]	.80	.23	.40	−.33	−.03
A	.75	.08	.35	−.42	.18
P	.66	.35	.26	−.34	.09
HOW I Feel About Things (HIF)					
W		.25	.27	−.37	−.14
A		.07	.31	−.32	.14
P		.32	.21	−.33	.08
Things I Like (Likes)					
W			.13	−.11	−.03
A			.21	−.10	.05
P			.04	−.34	.04
How Would You Finish It (S-C)					
W				−.12	−.07
A				−.14	.15
P				−.08	.18
Rogers Test of Personality					
W					.11
A					−.14
P					.20

[a] Abbreviations for *Well* Adjusted, *Average* Adjusted, and *Poorly* Adjusted Groups reading from top to bottom.

[b] The Rogers Test is scored in inverted order as compared to other tests.

Inspection of Table 33 containing these data reveals several striking facts. First, the California Test of Personality and the How I Feel About Things Test are highly related. Individuals who score high on one tend to score high on the other. This is independent of the adjustment group to which they belong. This is to say that, in regard to the criterion groups, those who are false positives or false negatives on the California will be in the same category on the other test. This is not true on the other tests for the correlations are moderate to no relationship at all.

The second most striking fact is that there are no other relationships that even begin to approach the California and How I Feel About Things similarity. This is of considerable theoretical importance, for the other tests significantly relate to the adjustment criterion. The extreme difference in types of items might be of considerable importance in eventually building a more adequate instrument.

Last, the results in terms of the school and clinical use of the tests is, in a negative sense, quite clear. In all but one case mentioned above the interpretation of high scores on one test meaning the same as high scores on another test is equivocal. For instance, those individuals regardless of adjustment level who have a high score on the Group Conformity Rating may have a high or low score on the California or on the Rogers. Indications of maladjustment on the one is not an indication of maladjustment on either of the others.

Summary

The purpose of this investigation is to analyze the adequacy or validity of six tests of adjustment for children. The results are tabulated in terms of statistical significance of mean differences and amount of overlap of the distributions. Five of the tests show significant mean

differences between the three groups. Such findings indicate a certain degree of validity of the tests. Such validity is probably most adequate for research purposes. When the effect of intelligence, reading achievement, and parental occupation are held constant, the *F* value usually drops slightly but generally not enough to lower the significance below the 5% level. Distributions of test scores reflecting amount of overlap indicate small practical differences between the groups. In some instances, the test results in combination with other data may be used with beneficial results. Use of the results from these tests alone for individual prediction, or use of any of the tests as a screening device, is extremely hazardous.

Discussion and Implications

The original purpose of this investigation, as stated previously, is to inquire into the adequacy of several personality and adjustment tests for children. Adequacy might be considered in terms of valid prediction in such specific tasks as screening, contributions to case studies, and research. The findings presented in the previous chapter show considerable statistical significance supporting the value of the tests in terms of a basic validity. Computation of percentages of overlap show the hazards in making predictions or statements about the individual case from these same test results.

The present section discusses a logical problem, that of base rates, which accents the interpretation that the present findings are negative. The remainder of this section analyzes three possible causes for the limited validity of the tests: (*a*) the individuals selected at the several levels of adjustment do not really represent such groups or levels, (*b*) the tests themselves are inadequate and do not discriminate because they were inadequately constructed, (*c*) the conceptual framework of level of adjustment indicated by single scores is inadequate and does not fit the actual dimensions of the situation. Such a discussion is not an attempt to rationalize away the findings but to indicate more clearly the implications of the findings in the areas of theory and measurement.

The Base Rate Problem

Meehl and Rosen (33) recently opened a very crucial problem related to investigations in the efficacy of tests as predictors or contributors to diagnosis. The most succinct definition of the base rate problem is the title of their article, "Antecedent Probability and the Efficiency of Psychometric Signs, Patterns, or Cutting Scores." They state: "Since diagnostic and prognostic statements can often be made with a high degree of accuracy purely on the basis of actuarial or experience tables (referred to hereinafter as base rates), a psychometric device, to be efficient, must make possible a greater number of correct decisions than could be made in terms of the base rates alone" (33, p. 194).

They discuss an example wherein a test with a certain cutoff score identified 55% of the valid positives and only 19% of the false positives. If such a test is used in the initial screening of a large group from which 5% is the amount of failure, the instrument would be 79.7% effective; however, if everyone is predicted for good adjustment, one would be 95% accurate. The latter figure occurs because the rate of incidence based on previous experience, records, or actuarial tables is only 5%. In such a situation a test has a difficult time contributing toward higher accuracy of prediction or judgment.

In the present investigation of adjustment tests for children, the instruments have yielded reliable differences in mean scores; however, due to the overlap of the distributions the percentage of false positives is excessively high. This is especially true of the average adjustment group. When this is complicated by the fact that the incidence of maladjustment in children is defined as only a small percentage of the total population, 14% in the present study, the use of any of these tests as a preliminary screening device, say in the fall of the year, is hazardous. The base rates of true positives are so low that the amount of error introduced by such testing would only confound the problem.

Following logically from these points, Meehl and Rosen discuss a way in which test signs or cutoff scores can aid in prediction. They refer to the fact that different populations have different percentages of incidence of valid positives. Knowledge of these varying rates or incidences in different groups to be examined is highly important. Such situations are quite typical when case studies are involved.

Consider the following example. Assume that the incidence of maladjustment in an elementary school class based on teacher judgment varies from 5% to 15% of the group. Identifying such children poses one predictive problem. But maladjustment of a pupil in such a class is dependent on problems—those within the individual, those related to the teacher, those related to the group, and those related to the interaction of the three factors. If an evaluation

of case records of referrals to the school psychologist indicates that 50% of the referred children ultimately receive psychiatric care privately or in an agency, the prediction problem has shifted. The base rate of maladjustment (maladjustment now defined as pupils receiving psychiatric care) is now 50 times in 100 rather than our previous 5 to 15 times in 100. This is quite a different prediction problem. The test which could pick up 55% valid positives and only 19% false positives is now an important aid in determining whether the problem lies within the child or within the other variables in the adjustment equation.

However, recall that the best single test, while identifying 50% of the true positives, indicated 15% of false positives from the well adjusted group and 27% false positives from the average adjusted group. In short, the base rate problem, the incidence of the disability in the population, accents the earlier interpretation. For screening purposes these tests are decidedly inadequate. For purposes of case study analysis, where the base rate of maladjustment is nearer 50%, the tests may contribute to the clinician's understanding, albeit a very limited amount.

Rationale for the Findings

At this point one is quite apt to question the adequacy of the three groups or levels of adjustment. Are these groups, in fact, different? The only defense is to review the literature, the methods used in selecting the children, and independent validity estimates presented earlier.

Whether one does or does not accept the groups as different, a second important consideration should still be raised. In the monograph, *The Prediction of Personal Adjustment*, Horst (22) outlines and discusses the logical and practical problems involved in the area of test construction. Basically, the six tests of personality and adjustment are measurements of an aspect of an individual's behavior which can be used to predict behavior. Even though Horst's analysis is in terms of prediction and resembles the discussion of predictive validity in the APA test construction manual, most of the observations are applicable to instruments dealing with concurrent validity. The general construction of the tests can be reviewed briefly in terms of Horst's analysis. This allows one to see assets and liabilities, at least in terms of one approach to measurement. Bias enters at this point, for Horst presents a consistent "empirical" point of view.

Chapter I of his monograph gives the overview of his approach: "The basic steps in the construction and application of predictive instruments are the same for all areas of human activity" (22, p. 4). The five steps which he outlines for the construction of a predictive instrument, in paraphrased form, are: (*a*) an index or measure of success selected—this is the criterion; (*b*) data on background factors, abilities, and personality are assembled and their relation to the criterion is determined; those that show a significant relationship become predictive items; (*c*) these items are combined; (*d*) selected items are then crossvalidated or reapplied to another sample; (*e*) if the relationship is maintained, then the predictive factors or instrument can be reapplied to the population with a known degree of success. The above procedures are easily interpreted in building an adjustment test.

Step one, the problem of obtaining a measure or index of success, is quite difficult. Horst comments on some of the hazards and difficulties involved: "First what constitutes success in an activity and second should a single or multiple criterion of success be used" (22, p. 23). In terms of criteria, how did the test authors define a maladjusted and an adjusted child? Did they use ratings by parents or teachers, were psychiatric opinions the measure, or were objective indices of delinquencies, and presence of nervous symptoms, the criterion measure? Or in some cases was there an objective criterion at all?

In this basic first step, considerable variability occurs in the construction of the six tests. The authors of the California test did not specifically have groups of adjusted and maladjusted children for a criterion. They selected behavior situations which logically should reflect maladjustment and wrote items which they felt logically would separate individuals with varying degrees of adjustment. Rogers had psychiatric ratings as his criterion measure and compared his test responses to these psychiatric ratings although his sample was quite small. The Nobles County investigators used as their criterion a summary of scores of 13 tests, each with an unknown validity. They assumed that these measures each possessed a limited degree of validity, and then they analyzed out the items most significantly related to this standard. Rosenzweig did not begin with groups of extra, intro-, or impunitive children. The GCR of his test was determined by statistical frequency of responses in the norm group.

The second step, the data on factors related to the criterion, is equally important. Much of the broad range of possible predictive correlates which Horst mentions are not possible in the construction of a group adjustment test. One is limited to verbal response, whether yes-no, like-dislike, more than-less than. Most important at this point is to obtain a wide variety of stimuli to which responses can be made. If verbal stimuli are used they should sample a wide range of factors: for children these might be attitudes toward the universal situations of

family, parents, siblings, friends, play activities, and school. The factors might also be fears, ideals, and goals. They may relate to everyday activities, eating, sleeping, playing, and working. The range of variables, even on a verbal level, is great, and prediction is enhanced by allowing for breadth of initial possibilities. The authors of all six tests give more adequate consideration to this problem. The breadth and variety of items and item types are considerable.

The combination of items is desirable for at least three reasons, according to Horst: (*a*) to obtain the most fundamental and generally applicable variables, (*b*) to reduce computational labor, and (*c*) to increase the reliability of prediction for samples other than the first. This problem seems only partially handled by the authors of the six tests. The California test, for instance, is long, and the two subtests show a high degree of correlation, indicating some inadequacies of selecting test variables and combining them. Tests such as the How Would You Finish It? and the Rosenzweig P-F Study when analyzed into components indicate the frequencies of the variables are too low to give basis for generalization of findings. Here the combinations are too severe and the tests are too short.

Step four in Horst's analysis concerns the problem of crossvalidation. Although much neglected by test authors, this is basically a simple point. For example, an investigator who has very carefully gone through the preceding steps might be faced logically with the following results. An adequate criterion of adjustment may have been selected; the original item pool may have numbered a thousand items. Out of this the investigator may have desired a 50-item test. On a chance basis, 5 out of every 100 items would be expected to be significant. By selecting these items his entire 50-item test could be composed of chance results. Crossvalidation eliminates this possibility by reapplying the selected test items to a new sample and determining the relationships to the criterion. After one, or more, such crossvalidation and selection of items the final instrument is developed.

This problem in construction has received limited attention by the test authors. The California Test of Personality's nonempirical construction in step one precluded any crossvalidation analysis. Rogers developed his scoring key on 22 of his cases and analyzed the test's validity on the whole 52 (including the 22). The P-F Study received no crossvalidation analysis. The Nobles County tests constitute a different category in that they are unpublished and still in the process of being developed. Crossvalidation in terms of concurrent and predictive validity is in process. However, in terms of their present status one comment seems warranted. In selecting the items for the How I Feel About Things the authors had approximately twice the number of items indicating statistically significant differences beyond the .05 level between the extreme groups used for their item analysis. Rather than reapplying this whole group of items to a new sample they selected the 40 most significant items and assumed that these would be the most significant with any other sample.

Step five is the application of the predictive instrument to the population and the determination of its efficiency in selection. At this point the other uses of tests, particularly the case study situation where the test uses additional information to be combined with other information, enter the discussion. However, such a difference is one of degree rather than kind and Horst's analysis applies nonetheless. Although Horst does not discuss this in these terms, this is the problem of false positives and false negatives. In the present discussion it is the percentage of the adjusted group labeled maladjusted by the instrument, and the percentage of maladjusted labeled adjusted by the instrument. Such continuing analysis has been followed only in a limited way by the test authors. The Nobles County tests are still unpublished and are being continually evaluated. The California test and the Rosenzweig P-F Study receive attention by investigators, but the test authors have not continued this type of research on any broad scale. The Rogers test has received little analysis beyond the original monograph.

In short, each of the six tests can be given an a priori estimate of adequacy in terms of Horst's conception of test construction. Only negligible attempts have been made to identify and to apply an adequate criterion of adjustment. Variety of items to be used has been fairly well handled. The combination of items has followed largely a nonempirical procedure. Crossvalidation is an unknown or unapplied procedure for several of the tests. Continuing analysis in the application of the tests to a variety of groups and populations has occurred only sparingly. The present investigator's opinion is that all of the tests leave much to be desired in their development and construction. The important unanswered question is: How much of the overlap of the distributions present in the findings could be eliminated by a properly constructed test?

Although this discussion of the results indicates that the tests are not adequate for individual prediction of adjustment, that there are a number of serious faults in the construction procedures of each of the tests, and that the

criterion groups are really different in adjustment, a further problem must be considered. The present writer, at the beginning of the design formulation, was struck by several facts: (*a*) the majority of the group type of children's personality and adjustment tests arithmetically sum all items and subtests to give a single score reflecting the individual's "adjustment"; (*b*) sociometric indices, also giving a set of scores on a single continuum, show marked relationships with adjustment; (*c*) teacher nominations have frequently been used to select children who are adjusting well or adjusting poorly; (*d*) several conceptions of adjustment, such as Havighurst's and Symonds', lead to the idea that there is a measurable state of the individual which varies depending on the individual's adjustment. These influences led to the conception and execution of the problem as has been done in this investigation. However, continued consideration of the problem plus the interesting results raised new issues and brought a re-evaluation of earlier conclusions, conclusions which had become a part of the design.

The empirical findings of the present investigation indicate that the tests are discriminating significantly, statistically if not practically, between the groups. Apparently a commonality within each group exists. The tests were built with varying approaches which raises doubts that an artifact of test construction produced the differences. Perhaps the dual criterion procedures, teacher and peer nominations, creates a commonality within each level of adjustment. The consistent statistical difference between the groups on the tests raises the important question —if these tests had been more carefully constructed empirically would they yield more practically valid results?

However, some doubt exists in such an interpretation, for some of the literature suggests diversity of types of persons appearing within each level. The tests with a single score indicating adjustment may never be more than barely adequate because of this diversity. For instance, how does such a test account for two maladjusted individuals—one with a very high degree of intropunitive guilt and another with very low or no guilt? The yes-no question when scored on a single key can only produce results which cancel each other. This is true of the hysteroid individual who superficially likes all activities on a like-dislike test and the negative child who dislikes every activity, in the environment. In rereading Northway's study (36) one finds that recessive children, socially uninterested children, and socially aggressive children comprise the low group. Bonney (9) and Jennings (25) and many others find similar diversity. Rereading the data on teacher nominations indicates the same phenomenon—diversity in behavior, symptoms, and personality structure of the individuals at the different levels. This is the Jenkins and Glickman (24) finding also.

In retrospect, such observations appear self-evident, natural, and without need of argument; yet thousands of dollars are spent buying the most widely used children's test, the California Test of Personality, which sums to a single score. The two subtests correlate highly, and the 12 scales are too unreliable for any kind of individual prediction. Thousands of man-hours are spent developing new instruments such as the How I Feel About Things Test, which also sums to give a single score.

Perhaps the most crucial problem at the present time is not building new tests nor analyzing old ones but attacking the criterion problem. Cattell (13) has a 12-factor conception of personality. Jenkins and Glickman (24) have a five-factor group of maladjusted children. Patterson[9] presents another factor analysis of children referred to a guidance clinic. His groups only partially coincide with Jenkins' and Glickman's. Havighurst (21) has a number of tasks which warrant individual analysis.

The present investigator leans heavily toward the building of a measuring instrument with multiple scales based on a conception of adjustment as a multiple problem. Whether this be a yes-no type test with a multiphasic-like profile or a picture test with a multiple scoring system such as the Rosenzweig P-F Study cannot be determined at the present time.

In discussing the findings of this investigation cognizance is taken of the "base rate problem." Such analysis sharply shows the limitations of using the six tests in certain kinds of situations and applications. In discussing the theme "why didn't the tests show more positive results in discriminating between the adjustment groups," several factors were considered. The experimental procedure was re-evaluated for errors and weakness, and points were indicated where more desirable steps could have been taken. The construction of the tests, when analyzed in terms of one prominent approach, show several grievous errors. The rationale of homogeneity of verbal behavior, symptoms, or personality within the adjusted or maladjusted was analyzed, and serious questions were raised. Implications for future research were drawn.

Summary and Conclusions

The adequacy of several personality and adjustment tests is the subject of this investigation. Selection of the tests de-

[9] Patterson, G. Personal communication, 1955.

pended on possibility of group administration, previously established or potential validity, wide usage, and the representation of a variety of item types. The selected tests seem to be as satisfactory a group as can be obtained from the personality and adjustment tests currently available. Adequacy has been referred to as the usefulness or validity of the tests as screening devices, as contributors of information in case studies of individual subjects, and as aids in research problems. The general approach used to analyze the tests falls under the rubric "concurrent validity."

Summary of the Methodology

The population sampled in this investigation is the public school children of St. Paul, Minnesota. The results should not be generalized to nonurban groups. The community contains an extremely large nonpublic school enrollment—40% of the elementary school pupils. This introduces an atypical factor of unknown proportion. Fewer than half the public elementary schools were used in this investigation because of sixth grades being platoon or being split with a fifth or seventh grade. No serious bias appeared through such a selection. The school system provides a special class program for pupils with mental and physical, both sensory and motor, handicaps; however, pupils with high ability, specific learning disabilities, and various emotional and adjustment problems remain in regular class.

Each of the 37 sixth-grade teachers nominated several boys in his class who he thought were very well adjusted and very poorly adjusted. In this manner the boys in the class were automatically split into three groups. All the children in the class contributed nominations of best friends and classmates with whom they did not get along. A hierarchy of acceptance was established. This was arbitrarily broken into thirds. A dual criterion for selection was invoked: to become a member of the well adjusted group a pupil must be nominated as well adjusted by the teacher and be in the top third of the peer acceptance hierarchy. Similarly both criteria were used in establishing the average and poorly adjusted groups. Any discrepancy in opinion or rating eliminated the pupil. A number of pupils were eliminated because of poor reading. Several pupils were lost due to absence and moving.

The sample was described in terms of age, intelligence, reading achievement, and level of parental occupation. Significant differences occurred on all variables except age. The relationship was positive between level of adjustment and brightness, achievement, and high socioeconomic status.

Independent estimates of the validity of the criterion groups were made. Differences occurred in referrals to schools social workers and the child guidance clinic; both supported the selection techniques. Membership on the school police patrol is related to adjustment. The frequency of arrests is indirectly related to adjustment. All four estimates support the basic rationale.

Summary of the Results

The results of the tests were analyzed in terms of mean differences between the groups, the analysis of variance being the statistical technique. Mean differences with intelligence, reading achievement, and level of parental occupation held constant were handled with the analysis of covariance technique. Tables of percentages of overlap further indicated the nature of the distribution differences.

The three adjustment groups showed highly significant statistical differences in mean scores on the California Test of Personality. The groups remained significant when the analysis of covariance technique was used with the previously cited three variables. The percentages of distribution overlap computed using the results were quite high, especially in regard to the average adjusted group. Although the California test proved as adequate as any of the others, it can be stated for this test as well as the others that using the test for screening purposes in a typical sixth-grade elementary school classroom is not satisfactory. The percentages of false positives and false negatives is very high. The test can make a contribution, to a certain degree, to case study analysis. The test is most applicable for research purposes when the purpose is to obtain rough comparisons between groups. The subtests, Personal Adjustment and Social Adjustment, are similar in mean and standard deviation; they are highly correlated.

The How I Feel About Things test also indicated statistically significant differences. Holding constant IQ, reading achievement, and socioeconomic status reduced the differences, but they remain significant. The overlapping of the distributions was quite marked. An interpretation comparable to that given the applicability of the California test holds for this test also.

The How Would You Finish It? test showed borderline statistical significance between the groups on mean scores. The Things I Like To Do test was not significantly different between the adjustment groups in means. Slight variations occurred on both tests when intelligence, reading achievement, and level of parental occupation were held constant. The probability figures fluctuated slightly above or below the 5% level. The distributions for the three groups on both tests overlapped almost completely. These tests do not indicate adjustment status of school children as adjustment has been defined in this investigation.

The Rogers Test of Personality Adjustment significantly differentiates the three adjustment groups. As with the other tests the statistical differences are overshadowed by the overlapping nature of the distributions. The same limitations in applicability discussed in relation to the California test hold for the Rogers test also.

The Group Conformity Rating of the Rosenzweig P-F Study shows statistically significant differences between the adjustment groups. These differences remain significant when the three descriptive variables are held constant. The distributions of the three groups overlap markedly. There are no significant trend changes in the GCR from the first to the last half of the blank. Only 3 out of 15 variables representing direction of aggression and type of reaction are significantly different between the adjustment groups. The chi-square technique indicates that the significance is at the borderline 5% level in each case.

In conclusion, it is important to repeat the limits in applicability of the test results. Within the population limits discussed earlier (urban community and public school setting) and within the procedural limitations (criterion groups based on teacher and pupil nominations), the results can be applied as follows. None of the tests discriminates well enough to be used as a selection instrument for maladjusted pupils in the usual classroom. Several of the tests indicate group differences extensive enough to allow for a limited amount of applicability

when the test data are combined with other information in studying an individual case. At least four of the tests relate well enough to adjustment status so that their most appropriate use at the present time is in research.

Summary of the Discussion

A logical problem, in the base rate of incidence of the factor to be discriminated by a test, was discussed in relation to the present findings. The inadvisability of using any of these tests as a screening instrument when the experience table indicates only 5% to 15% of the population is maladjusted was observed. The changing base rates with a different subpopulation, i.e., referrals to the school psychologist, indicated that the tests could be used for limited practical purposes.

In analyzing the possible factors underlying such basically negative, or at best mildly positive, findings, several problems were considered. These are the nature of the criterion groups, the manner in which the tests were constructed, and the nature of the concept of adjustment. The procedures used in devising the tests were found to be very inadequate in terms of one approach to measurement and test construction. The inadequacies were found to range all the way from initial analysis of the criterion to cross-validation and continuing analysis of the application of the instrument.

The concept of adjustment was reconsidered. Emphasis was placed on the probable variety of behaviors, symptoms, and verbalization of individuals at any of the three adjustment levels. Illustrations indicating some of the inadequate aspects of an adjustment test with a single summary score were given. Research oriented more toward the basic dimensions of adjustment was suggested. A test with several scales or a multiple scoring system seems most likely to fit the dimensions of the situation.

General Conclusion

This study approached the problem of the adequacy of several adjustment tests for children. The conceptual framework was that of concurrent validity. Limitations occur in the generalizability of the findings because of the population, procedures, and resulting sample. These dimensions have been stated carefully at several points in the report. The nature of the findings indicate additional limitations; these limitations are in terms of applicability of the test results.

REFERENCES

1. Ackerson, L. *Children's behavior problems.* Chicago: Univer. of Chicago Press, 1931.
2. American Psychological Association, American Educational Research Association, & National Council on Measurements Used in Education. Joint Committee. Technical recommendations for psychological tests and diagnostic techniques. *Psychol. Bull.*, 1954, **51**, 201–238.
3. Ammons, R. B. *Manual, Full Range Picture Vocabularly Test.* Beverly Hills: Western Psychol. Services, 1948.
4. Anderson, J. E. Developmental level and adjustment in children and youth. Paper read at the Psychology Section of the British Ass. for the Advancement of Sci., Edinburgh, August, 1951.
5. Anderson, J. E. Nobles County every-child survey. Paper read at APA, New York, September, 1952.
6. Anderson, J. E. The relation of attitude to adjustment. *Education*, 1952, **73**, 1–9.
7. Austin, M. C., & Thompson, G. C. Children's friendships: a study of the bases on which children select and reject their best friends. *J. educ. Psychol.*, 1948, **39**, 101–116.
8. Bonney, M. E. Personality traits of socially successful and socially unsuccessful chil-

dren. *J. educ. Psychol.*, 1943, **34**, 449–472.
9. BONNEY, M. E. Popular and unpopular children, a sociometric study. *Sociometry Monogr.*, 1947, No. 9.
10. BOYNTON, P. S., & MCGAW, BONNIE H. The characteristics of problem children. *J. juv. Res.*, 1934, **18**, 215–222.
11. BUROS, O. K. (Ed.) *The fourth mental measurements yearbook*. New Brunswick: Rutgers Univer. Press, 1953.
12. CAMPBELL, NELLIE M. The elementary school teacher's treatment of classroom behavior problems. *Teach. Coll. Contrib. Educ.*, 1935, No. 668.
13. CATTELL, R. B. *Personality*. New York: McGraw-Hill, 1950.
14. CLARK, E. J. Teacher reactions toward objectionable behavior. *Elem. Sch. J.*, 1951, **51**, 446–449.
15. ELLIS, A. The validity of personality questionnaires. *Psychol. Bull.*, 1946, **43**, 385–440.
16. ELLIS, A. Recent research with personality inventories. *J. consult. Psychol.*, 1953, **17**, 45–49.
17. FORLANO, G., & WRIGHTSTONE, J. W. Sociometric and self descriptive techniques in the appraisal of pupil adjustment. *Sociometry*, 1951, 340–350.
18. FOX, W. H., & SEGAL, D. The validity of the choice of friends method of measuring social adjustment. *J. educ. Res.*, 1954, **47**, 389–394.
19. FRANKEL, E. B., & POTASHIN, R. A survey of sociometric and pre-sociometric literature on friendship and social acceptance among children. *Sociometry*, 1944, **7**, 422–431.
20. GATES, *A. I. Manual of direction for the Gates Advanced Primary Reading Test*. New York: Bureau of Publications, Teachers Coll., Columbia Univer., 1943.
21. HAVIGHURST, R. J. *Human development and education*. New York: Longmans, Green, 1953.
22. HORST, P. (Ed.) *The prediction of personal adjustment. Soc. Sci. Res. Coun. Bull.*, 1941, No. 48, 1–156.
23. HUNT, J. MCV., & SOLOMON, R. L. The stability and some correlates of group status in a summer-camp group of young boys. *Amer. J. Psychol.*, 1942, **55**, 33–45.
24. JENKINS, R. L., & GLICKMAN, S. Common syndromes in child psychiatry. *Amer. J. Orthopsychiat.*, 1946, **16**, 244–261.
25. JENNINGS, HELEN H. *Leadership and isolation*. (2nd ed.) New York: Longmans, Green, 1950.
26. KAPLAN, L. The annoyances of elementary school teachers. *J. educ. Res.*, 1952, **45**, 649–665.
27. KEARNY, N. C. et al., Final report on the opinionnaire study of the over-all elementary curriculum committee to the curriculum steering committee. Curriculum Bull. No. 44. St. Paul Public Schools, 1952. (Mimeographed)
28. KUHLEN, R. G., & LEE, B. J. Personality characteristics and social acceptability in adolescence. *J. educ. Psychol.*, 1943, **34**, 321–340.
29. LAYCOCK, S. R. Teachers' reactions to maladjustment of school children. *Brit. J. educ. Psychol.*, 1934, **4**, 11–29.
30. LOUTTIT, C. M. Test review. In O. K. Buros (Ed.), *The 1940 mental measurements yearbook*. Highland Park, N.J.: Mental Measurements Yearbook, 1941. P. 94.
31. MCCLURE, W. E. Characteristics of problem children based on judgments of teachers. *J. juv. Res.*, 1929, **13**, 124–140.
32. MACFARLANE, J. W., HONZIK, M. P., & DAVIS, M. H. Reputation differences among young school children. *J. educ. Psychol.*, 1937, **28**, 161–175.
33. MEEHL, P. E., & ROSEN, A. Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychol. Bull.*, 1955, **52**, 194–216.
34. MITCHELL, J. C. Study of teachers' and of mental hygienists' ratings of certain behavior problems of children in 1927 and in 1940. *J. educ. Res.*, 1942, **36**, 292–307.
35. NEWSLETTER, W. I., FELDSTEIN, J. J., & NEWCOMB, T. M. *Group adjustment*. Cleveland: Western Reserve Univer. Press, 1938.
36. NORTHWAY, MARY L. Appraisal of the social development of children at a summer camp. *Univer. of Toronto Studies, Psychol. Series*, 1940. Vol. 5, No. 1.
37. NORTHWAY, MARY L. Outsiders, a study of the personality factors of children least acceptable to their age mates. *Sociometry*, 1944, **7**, 10–25.
38. OTIS, A. S. *Manual for the Otis Quick Scoring Mental Ability Tests*. Yonkers, N.Y.: World Book Co., 1954.
39. PECK, LEIGH. Teachers' reports of the problems of unadjusted school children. *J. educ. Psychol.*, 1935, **26**, 123–138.
40. PEPINSKY, P. N. The meaning of "validity" and "reliability" as applied to sociometric tests. *Educ. psychol. Measmt*, 1949, **9**, 39–49.
41. ROGERS, C. R. *Manual of directions, A Test of Personality Adjustment*. New York: Association Press, 1931.
42. ROGERS, C. R. Measuring personality adjustment in children nine to thirteen years of age. *Teach. Coll. Contrib. Educ.*, 1931, No. 458.

43. ROSENZWEIG, S., FLEMING, EDITH, & ROSENZWEIG, LOUISE. The children's form of the Rosenzweig P-F Study. *J. Psychol.*, 1948, **26**, 141–191.
44. SHAFFER, L. F. Test review. In O. K. Buros (Ed.), *The third mental measurement yearbook*. New Brunswick: Rutgers Univer. Press, 1949.
45. SMITH, LOUIS M. A validity study of six personality and adjustment tests for children. Unpublished doctoral dissertation, Univer. of Minnesota, 1955.
46. SNYDER, LOUISE M. The problem in the Jersey City elementary schools. *J. educ. Sociol.*, 1934, **7**, 343–352.
47. SPARKS, J. N. Teachers' attitudes toward the behavior problems of children. *J. educ. Psychol.*, 1952, **43**, 284–291.
48. STOUFFER, G. A. W. Behavior problems of children as viewed by teachers and mental hygienists. *Ment. Hyg., N.Y.*, 1952, **36**, 271–285.
49. SYMONDS, P. M. *The ego and the self*. New York: Appleton-Century-Crofts, 1951.
50. TAGIURI, R. Relational analysis: an extension of sociometric method with emphasis upon social perception. *Sociometry*, 1952, **15**, 91–104.
51. THOMPSON, C. E. The attitudes of various groups toward behavior problems of children. *J. abnorm. soc. Psychol.*, 1940, **35**, 120–125.
52. THORPE, L. P., CLARK, W. W., and TIEGS, E. W. *Manual, California Test of Personality*. Los Angeles: California Test Bureau, 1953.
53. UNIV. OF MINNESOTA, INSTITUTE OF CHILD WELFARE. Minnesota scale for paternal occupations. Minneapolis: Univer. of Minnesota. (leaflet), Rev. 1952.
54. WATSON, G. B. A critical note on two attitude studies. *Ment. Hyg., N.Y.*, 1933, **17**, 59–64.
55. WICKMAN, E. K. *Children's behavior and teachers' attitudes*. New York: The Commonwealth Fund, 1928.
56. YOUNG, L. L., & COOPER, D. H. Some factors associated with popularity. *J. educ. Psychol.*, 1944, **35**, 513–535.
57. YOURMAN, J. Children identified by their teachers as problems. *J. educ. Sociol.*, 1932, **5**, 334–344.

(Accepted for publication February 13, 1957)

Appendix

TEACHER NOMINATION FORM
RESEARCH PROJECT IN ADJUSTMENT

This is part of a project in the identification and testing of adjustment in children. As was pointed out in the earlier letter or personal contact your name will not be associated with any individual pupils. All the data from the several schools and numerous teachers will be grouped for anonymity of all participants both teacher and pupil. Once again thanks for your cooperation.

1. List below the three to five *boys* in your class who are the most well-adjusted. These children get along very well with their classmates, they like the other kids and have many friends, they are happy, secure, and self-confident with no more than normal worries, they don't cry or lose their tempers easily, they are not "goody-good" but engage in normal child activity; they are *not* bullies, they don't fight excessively, nor do they shy away from other children; they do not fear new situations. If their present growth and development continues they should be well-adjusted, happy, useful adults. Select on the basis of total impression.

Boys

1.
2.
3.
4.
5.

2. List the *boys* in your class, three to five, who are the most severely emotionally disturbed. Among the things they do may be one or several of the following: they do not get along well with the other children or the teachers, they have few friends, they cry easily, worry a lot, are apt to lose their tempers, are unhappy; they fight a lot and bully other kids; they daydream excessively, don't join in activities and don't play well, and they act odd. Some children may fit one description or another but list the ones who you think are the most seriously maladjusted and emotionally disturbed, some of whom might someday need psychiatric help if present development isn't altered. Again, select on the basis of total impression.

Boys

1.
2.
3.
4.
5.

3. List the one or two *boys* you haven't considered for the above categories because you don't know them well enough.

1.
2.

TABLE 34

Results of the Bartlett Test for Homogeneity of Variances of the Three Adjustment Groups for Each of the Analysis of Variance Problems

Table No.	Variable	Chi square	p	Assumption
6	Age	18.43	$<.01$	Unsatisfied
6	Intelligence	.59	$>.05$	Satisfied
6	Reading Achievement	.26	$>.05$	Satisfied
6	Parental Occupation	1.11	$>.05$	Satisfied
13	C.T.P. Total Score	1.62	$>.05$	Satisfied
13	C.T.P. PA Subtest	2.12	$>.05$	Satisfied
13	C.T.P. SA Subtest	2.53	$>.05$	Satisfied
17	How Would You Finish It?	1.60	$>.05$	Satisfied
21	How I Feel About Things Total Score	1.08	$>.05$	Satisfied
21	How I Feel About Things Adjustment Subtest	.57	$>.05$	Satisfied
21	How I Feel About Things Adjustment-Maturity	1.12	$<.05$	Satisfied
25	Things I Like To Do	.16	$>.05$	Satisfied
29	Rogers Full Scale	4.22	$>.05$	Satisfied
29	Rogers PI Scale	6.54	$>.01$	Unsatisfied
29	Rogers SM Scale	3.75	$>.05$	Satisfied
29	Rogers FM Scale	1.27	$>.05$	Satisfied
29	Rogers DD Scale	.64	$>.05$	Satisfied
34	Rosenzweig GCR	2.43	$>.05$	Satisfied

TABLE 35

RESULTS OF THE L_1 TEST OF HOMOGENEITY OF VARIANCES ABOUT THE REGRESSION LINE FOR EACH OF THE ANALYSIS OF COVARIANCE PROBLEMS

Table No.	Variables	L_1	p	Assumption
14	C.T.P. and IQ	.995	> .05	Satisfied
14	C.T.P. and Reading	.993	> .95	Satisfied
14	C.T.P. and SES	.984	> .05	Satisfied
18	How Finish and IQ	.993	> .05	Satisfied
18	How Finish and Reading	.995	> .05	Satisfied
18	How Finish and SES	.993	> .05	Satisfied
22	H.I.F. and IQ	.998	> .05	Satisfied
22	H.I.F. and Reading	.990	> .05	Satisfied
22	H.I.F. and SES	.991	> .05	Satisfied
26	Likes-Dislikes and IQ	.998	> .05	Satisfied
26	Likes-Dislikes and Reading	.999	> .05	Satisfied
26	Likes-Dislikes and SES	.999	> .05	Satisfied
30	Rogers and IQ	.970	> .05	Satisfied
30	Rogers and Reading	.980	> .05	Satisfied
30	Rogers and SES	.980	> .05	Satisfied
35	Rosenzweig GCR and IQ	.995	> .05	Satisfied
35	Rosenzweig GCR and Reading	.995	> .05	Satisfied
35	Rosenzweig GCR and SES	.997	> .05	Satisfied

DIRECTIONS FOR PEER NOMINATIONS

My name is Mr. Smith. Some of you know me, most of you probably do not. Today I'm here to ask your help and cooperation in a project. It is a survey of the friendships among boys and girls in the sixth grades from all over the city of St. Paul. First let me pass out a sheet of paper to all of you.

As I said before, this is a survey of friendships among boys and girls from all over the city. I want you to help me in kind of a secret ballot, a secret vote. All of you have a sheet of paper. On the top of the paper I want you to write down the names of the boys and girls in your room whom you like best of all. These are the ones you get along with best. You may put down all boys, all girls, or some of each. You may put down as many as you want, but most boys and girls usually like four or five best. Please don't sign your name. Go ahead.

Write both first and last names down. If you can't spell the last names be sure to put the first initial of the last name down. Do as best you can. If any of your best friends are absent today be sure and put their names down too.

Is most everyone about finished? Now I want you to draw a line across the middle of the page from one side to the other. (Demonstrate.) This is just to separate the names on top from the rest of the paper.

Now on the bottom of the page, under the line, we will continue our secret vote. This time I want you to put down the names of the boys and girls that you don't get along with very well. These children often have difficulties with the other children. You may select all boys, all girls, or some of each. Remember, this is confidential, no one but me will see what name you put on the paper. And be sure not to sign your name.

As you finish, and take a few more minutes if you need to, fold your paper in the middle so that the names are on the inside; I will collect them in this folder.

GEORGE BANTA COMPANY, INC., MENASHA, WISCONSIN